Don't stay local - extrapolation analytics for Dupire's local volatility

Peter Friz*, Stefan Gerhold
TU and WIAS Berlin, TU Wien

January 20, 2013

Abstract

A robust implementation of a Dupire type local volatility model is an important issue for every option trading floor. Typically, this (inverse) problem is solved in a two step procedure : (i) a smooth parametrization of the implied volatility surface; (ii) computation of the local volatility based on the resulting call price surface. Point (i), and in particular how to extrapolate the implied volatility in extreme strike regimes not seen in the market, has been the subject of numerous articles, starting with Lee (Math. Finance, 2004). In the present paper we give direct analytic insights into the asymptotic behavior of local volatility at extreme strikes.

1 A new formula for local volatility extrapolation

Volatility remains a key concept in modern quantitative finance. In particular, the Black-Scholes *implied volatility* surface $\sigma_{\mathrm{BS}} = \sigma_{\mathrm{BS}}(K,T)$ is the central object of any option trading desk, see e.g. [15]. On the quantitative and computational side, a smooth and arbitrage free parametrization of the implied volatility surface is a crucial step towards a robust implementation of a Dupire type local volatility model [11, 12]. Indeed, Dupire's formula

$$\sigma^2_{\mathrm{loc}}(K,T) = \frac{2\partial_T C}{K^2 \partial_{KK} C} \tag{1.1}$$

implies that any arbitrage free call price surface

$$C = C(K,T) = C_{\mathrm{BS}}(K,T;\sigma_{\mathrm{BS}}(K,T))$$

which arises from a (not necessarily Markovian) Itô diffusion is obtained from the one-factor ("Dupire's local vol") model

$$dS_t/S_t = \sigma_{\mathrm{loc}}(S_t,t)dW_t.$$

*partially supported by MATHEON

(Note that spot remains fixed in the present discussion, and that we work under the appropriate forward measure to avoid drift terms.) It is helpful to think of local volatility as a Markovian projection, term coined in [22], of a higher dimensional model (e.g. Heston); the first component then forms an Itô diffusion of the form

$$dS_t/S_t = \sigma_{\mathrm{stoch}}(t,\omega)dW_t.$$

Indeed, it is known (see e.g. [15] and the references therein) that $\sigma^2_{\mathrm{loc}}(K,T) = \mathbb{E}[\sigma^2_{\mathrm{stoch}} \,|S_T = K]$; in practice, this means that even for stochastic volatility models with fully explicit Markovian specification, sampling from the corresponding local volatility models requires substantial computational effort. (In particular, the singular conditioning requires Malliavin calculus techniques, as was pointed out e.g. in [17].)

The analysis of implied, local, and stochastic volatility and their interplay has been subject of countless works; a very small selection relevant to the present discussion is [1, 2, 4, 14, 16, 20]. Our contribution here is a formula ((1.6) below) that allows for approximation of $\sigma^2_{\mathrm{loc}}(K,T)$ when K is large (and similarly, K is small). The main ingredient to this formula is a known moment generating function (mgf) of the log-price (X_t) (under the pricing measure),

$$M(s,T) := \exp(m(s,T)) := \mathbb{E}\exp(sX_T),$$

assumed to be finite in some (maximal) interval $(s_-(T), s_+(T))$ with critical exponents s_- and s_+ defined as

$$s_-(T) := \inf\{s : M(s,T) < \infty\}, \qquad s_+(T) := \sup\{s : M(s,T) < \infty\}.$$

We also assume that call prices have sufficient regularity to make Dupire's formula (1.1) well-defined, and that the mgf blows up at the upper critical moment:

$$\lim_{s\uparrow s_+(T)} M(s,T) = \infty. \tag{1.2}$$

This holds, e.g., in the Heston model [18], with log-price $X_t = \log(S_t/S_0)$, where

$$\begin{aligned} dS_t &= S_t\sqrt{Y_t}dW_t, \qquad S_0 = s_0 > 0, \\ dV_t &= (a + bV_t)dt + c\sqrt{V_t}dZ_t, \qquad V_0 = v_0 > 0, \end{aligned}$$

with $a \geq 0$, $b \leq 0$, $c > 0$, and $d\langle W, Z\rangle_t = \rho dt$ with $\rho \in (-1,1)$. We will prove the following theorem, which is reminiscent of Lee's formula [20] for implied volatility.

Theorem 1. *In the Heston model with $\rho \leq 0$ (the relevant regime in practice), the following local volatility approximation holds:*[1]

$$\lim_{k\to\infty} \frac{\sigma^2_{\mathrm{loc}}(k,T)}{k} = \frac{2}{T\, s_+(s_+ - 1)R_1/R_2}, \tag{1.3}$$

[1] By a common abuse of notation, we write $\sigma^2_{\mathrm{loc}}(k,T)$ instead of $\sigma^2_{\mathrm{loc}}(e^k,T)$ when we wish to express the local vol as a function of log-strike k.

where $k = \log(K/S_0)$, $s_+ \equiv s_+(T)$ *and*

$$\begin{aligned} R_1 &= c^2 s_+(s_+-1)\left[c^2(2s_+-1) - 2\rho c(s_+\rho c + b)\right] \\ &\quad - 2(s_+\rho c + b)\left[c^2(2s_+-1) - 2\rho c(s_+\rho c + b)\right] \\ &\quad + 4\rho c\left[c^2 s_+(s_+-1) - (s_+\rho c + b)^2\right], \end{aligned} \tag{1.4}$$

$$R_2 = 2c^2 s_+(s_+-1)\left[c^2 s_+(s_+-1) - (s_+\rho c + b)^2\right]. \tag{1.5}$$

The origin of this result lies in the saddle point based approximation formula

$$\sigma^2_{\mathrm{loc}}(k,T) \approx \left. \frac{2\frac{\partial}{\partial T}m(s,T)}{s(s-1)} \right|_{s=\hat{s}(k,T)}, \tag{1.6}$$

where k denotes log-strike, and $\hat{s} = \hat{s}(k,T)$ is determined as solution of the equation

$$\frac{\partial}{\partial s} m(s,T) = k. \tag{1.7}$$

As such, our formula (1.6) is not restricted to the Heston model. As a trivial example, let us consider the generalized Black-Scholes model with time-dependent volatility,

$$dS_t = S_t\sqrt{v(t)}dW_t.$$

We find $m(s,T) = \frac{1}{2}s(s-1)\int_0^T v(t)dt$, and then, correctly,

$$\sigma^2_{\mathrm{loc}}(k,T) = v(T).$$

(In this example the evaluation of $\hat{s}(k,T)$ plays no role, since the fraction on the right hand side of (1.6) does not depend on it.)

In fact, we expect our approximation formula (1.6) to work whenever the saddle point method is applicable (also assuming that call prices are smooth enough to make (1.1) a well-defined quantity); the essence of the argument is given in Section 2. Of course, the ultimate justification of a saddle point approximation involves tail estimates which may present mathematical challenges (while easy to observe numerically); in the Heston case, we achieve this by a subtle application of ODE comparison results, applied to the underlying Riccati equations (cf. Appendix A), thus completing our proof of the above theorem. The asymptotic equivalence of (1.3) and (1.6) is discussed in Section 4.

Interestingly, even when the blow-up of the mgf is too slow to apply the saddle point method, the approximation formula (1.6) can give surprisingly accurate results. Our attempt to understand this phenomenon, besides numerical evidence in the variance gamma model for suitable parameters, discussed in Section (5), passes through Karamata's Tauberian theorem and is the content of Section 3. We have not pushed our investigations too far, however, since the meaning of Dupire's local volatility in the presence of jumps may be questioned (cf. our comment below on extension of Dupire's formula to jump settings.)

Various additional comments are in order.

1. Equation (1.7) is solvable for large k, since (1.2) implies

$$\lim_{s \uparrow s_+} \frac{\partial}{\partial s} m(s, T) = \infty.$$

2. If there is no blow-up, i.e. (1.2) does not hold, then (1.6) is typically incorrect. See Example 4 in Section 5 for some hints on how to handle such cases.

3. We have $\hat{s}(k, T) \uparrow s_+(T)$ as $k \to \infty$; hence, in models with moment explosion [1, 20], where $s_+(T) < \infty$, the denominator in (1.6) may be replaced by $s_+(T)(s_+(T) - 1)$. While this is correct to first order, it is often preferable to use (1.6) as it is, and to calculate $\hat{s}(k, T)$ by (numerically) solving (1.7). This tradeoff between simple formulas and numerical precision is illustrated in several examples in Sections 4 and 5. The comment applies in particular to the Heston model.

4. There is a version of our approximation formula (1.6) for *small* values of K (i.e. $K \downarrow 0$, or $k \downarrow -\infty$), which requires that the mgf blows up at the lower critical moment $s_-(T)$. If $k < 0$ and $|k|$ is large, equation (1.7) has a unique solution $\hat{s}_-(k, T) < 0$. Then the approximation (1.6) holds, if $\hat{s}$ is replaced by $\hat{s}_-$.

5. There are extensions of Dupire's work to jump diffusions and also pure jump models; the resulting "local" version of these models is studied in [3]. Local Lévy models were introduced earlier in [6]. In particular, Dupire's formula (which may be written as a PDE) becomes a PIDE which features an integral term involving the second derivative of C w.r.t. strike, times a kernel depending on K, integrated against all strikes in $(0, \infty)$. (The formula, which we need not reproduce here in full technical detail, appears in Theorem 1 of [3].)

   Another difficulty in the jump setting is the potential lack of immediate smoothing. For instance, the variance gamma model satisfies the above PIDE only in viscosity sense; in fact, call prices in the variance gamma model may not be twice differentiable in K for small times, as was noted in [9]. But for sufficiently large times our formula (1.6) works, see Example 3 in Section 5.

   We conclude that, in a general jump setting, Dupire's formula, as stated in (1.1), may be ill-defined; moreover, even if call prices are smooth enough to make the formula well-defined, it fails to recreate the correct marginals of the original price process.

6. Even so, given the industry practice of applying Dupire's formula to any given call price surface, we discuss in Section 5 what happens when applying (1.6) to jump models, if possible. Formula (1.6) simplifies in exponential Lévy models, which have the property that $m(s, T)$ is linear in T; thus, the numerator in (1.6) may be replaced by $2m(s, 1)$. In jump models,

we also expect $\sigma^2_{\mathrm{loc}}(k,T)$ to explode as $T \downarrow 0$, and we shall observe and quantify this blow-up in some examples below. There is potential practical value in that a Dupire local volatility surface, fitted to market data, may so be inspected for evidence of jump behavior (thereby questioning the use of Dupire's formula in the first place).

2 Saddle point asymptotics

As is well known [7], we can recover the call price C and the probability density $D(\cdot,T)$ of S_T by Laplace-Fourier inversion from the mgf:

$$C(K,T) = \frac{e^k}{2i\pi}\int_{-i\infty}^{i\infty} e^{-ks}\frac{M(s,T)}{s(s-1)}ds, \tag{2.1}$$

$$D(x,T) = \frac{1}{2i\pi}\int_{-i\infty}^{i\infty} e^{-(s+1)\log x}M(s,T)ds. \tag{2.2}$$

Now differentiate the call price under the integral sign:

$$\partial_T C(K,T) = \frac{e^k}{2i\pi}\int_{-i\infty}^{i\infty} \frac{\partial_T m(s,T)}{s(s-1)}e^{-ks}M(s,T)ds. \tag{2.3}$$

By Dupire's formula, we have

$$\sigma^2_{\mathrm{loc}}(k,T) = \frac{2\partial_T C(K)}{K^2 D(K,T)} = \frac{2\int_{-i\infty}^{i\infty}\frac{\partial_T m(s,T)}{s(s-1)}e^{-ks}M(s,T)ds}{\int_{-i\infty}^{i\infty}e^{-ks}M(s,T)ds}. \tag{2.4}$$

Both integrands in (2.4) have a singularity at $s = s_+$, since $M(s,T)$ gets infinite there. The singular behavior of $M(s,T)$ dominates the asymptotics of both integrals. The resulting asymptotic factor cancels, and only the contribution of $2\frac{\partial_T m(s,T)}{s(s-1)}$ remains. This is the idea behind (1.6).

To implement it, we analyze both integrals in (2.4) by a saddle point approximation [10]. If M features an exponential blow-up at the critical moment s_+, its validity can be justified rather universally. Examples include the Heston model, double exponential Lévy, and Black-Scholes. (Note that the critical moment is $s_+ = \infty$ for Black-Scholes.) If the saddle point method is not applicable (because of insufficient blow-up), different arguments are required; see the following section.

So let us proceed with the saddle point analysis of (2.4). For both integrals, we only use the factor $e^{-ks}M(s)$ to find the location of the (approximate) saddle point $\hat{s} = \hat{s}(k,T)$. The saddle point equation is (1.7), obtained by equating the derivative of $e^{-ks}M(s)$ to zero. We move the integration contour through the saddle point. Then, for large k, only a small part $|\Im(s)| \le h(k)$ of the contour, around the saddle point, matters asymptotically. (The choice of the function h depends on the singular expansion of M; see Section 4 for an example.) The integral can be approximated via a local expansion of the integrand. Let us carry

this out for the denominator of (2.4). (In the following formulas we write m'' for $\partial^2 m/\partial s^2$.)

$$\begin{aligned}
\int_{\hat{s}-i\infty}^{\hat{s}+i\infty} e^{-ks} M(s,T)ds &\sim \int_{\hat{s}-ih(k)}^{\hat{s}+ih(k)} e^{-ks} M(s,T)ds \\
&\sim \int_{\hat{s}-ih(k)}^{\hat{s}+ih(k)} \exp\left(-ks + m(\hat{s},T) + k(s-\hat{s}) + \tfrac{1}{2}m''(\hat{s},T)(s-\hat{s})^2\right) ds \\
&= e^{m(\hat{s},T)-k\hat{s}} \int_{\hat{s}-ih(k)}^{\hat{s}+ih(k)} \exp\left(\tfrac{1}{2}m''(\hat{s},T)(s-\hat{s})^2\right) ds. \qquad (2.5)
\end{aligned}$$

In the Taylor expansion of the exponent we have used the equation $m'(\hat{s},T) = k$. Now the crucial observation is that the numerator of (2.4) admits a similar approximation, where the only new ingredient is the factor $2\frac{\partial_T m(s,T)}{s(s-1)}$:

$$\begin{aligned}
&2\int_{-i\infty}^{i\infty} \frac{\partial_T m(s,T)}{s(s-1)} e^{-ks} M(s,T)ds \\
&\sim 2\int_{\hat{s}-ih(k)}^{\hat{s}+ih(k)} \frac{\partial_T m(s,T)}{s(s-1)} e^{-ks} M(s,T)ds \\
&\sim 2e^{m(\hat{s},T)-k\hat{s}} \int_{\hat{s}-ih(k)}^{\hat{s}+ih(k)} \frac{\partial_T m(\hat{s},T)}{\hat{s}(\hat{s}-1)} (1+o(1)) \exp\left(\tfrac{1}{2}m''(\hat{s},T)(s-\hat{s})^2\right) ds \\
&\sim 2\frac{\partial_T m(\hat{s},T)}{\hat{s}(\hat{s}-1)} e^{m(\hat{s},T)-k\hat{s}} \int_{\hat{s}-ih(k)}^{\hat{s}+ih(k)} \exp\left(\tfrac{1}{2}m''(\hat{s},T)(s-\hat{s})^2\right) ds. \qquad (2.6)
\end{aligned}$$

Dividing (2.6) by (2.5) concludes the derivation. Summarizing, we note that the asymptotics of $\sigma_{\mathrm{loc}}^2(k)$ are governed by the local expansions at $s = \hat{s}$ of the integrands in (2.4). The respective first terms of both expansions agree, and thus cancel, except for the factor (1.6).

3 Algebraic singularities and Karamata's theorem

The saddle point method is well suited to treat mgfs of exponential growth, such as $M(s,T) \approx \exp(1/(s_+ - s))$, but fails in cases of slower blow-up. To see how to analyze these, let us assume that the mgf M grows like a power at the (finite) critical moment:

$$M(s,T) \sim \frac{c_1}{(s_+ - s)^{c_2}}, \qquad s \uparrow s_+.$$

(The variance gamma model is a typical instance.) The quantities $c_1 = c_1(T) > 0$ and $c_2 = c_2(T) > 0$ are independent of s, but may be functions of maturity T. (In particular, we assume that c_2 does depend on T, which holds in

Lévy models.) Since

$$\frac{\partial}{\partial s} m(s,T) \sim \frac{c_2}{s_+ - s},$$

the saddle point $\hat{s}$ satisfies

$$\hat{s} \approx s_+ - \frac{c_2}{k}.$$

Inserting this into the time derivative

$$\frac{\partial}{\partial T} m(s,T) \sim \dot{c}_2(T) \log \frac{1}{s_+ - s}$$

of m yields

$$\left. \frac{2\frac{\partial}{\partial T} m\,(s,T)}{s\,(s-1)} \right|_{s=\hat{s}(k,T)} \sim \frac{2\dot{c}_2(T)\log k}{s_+(s_+ - 1)}. \tag{3.1}$$

To justify (1.6), we now have to argue that $\sigma^2_{\mathrm{loc}}(k,T)$ has the same asymptotics. Again, we use the representation (2.4). To put it briefly, the reason why the approach from the preceding section fails is that one cannot find a suitable $h(k)$. (Either the tails $|\Im(s)| > h(k)$ of the integrals are not negligible, or the local expansion is not uniformly valid.) But it is still true that the local behavior of the integrands near s_+ fully determines the asymptotics of the integrals in (2.4).

We write $f(\cdot,T)$ for the probability density of the log-price X_T. Note that the denominator in (2.4) equals $2i\pi f(k,T)$. The (one-sided!) Laplace transform of $k \mapsto e^{s_+ k} f(k,T)$ satisfies

$$\int_0^\infty e^{-sk} e^{s_+ k} f(k,T)dk \sim c_1 s^{c_2}, \qquad s \downarrow 0. \tag{3.2}$$

This follows from

$$M(s,T) = \int_{-\infty}^0 e^{sk} f(k,T)dk + \int_0^\infty e^{sk} f(k,T)dk \sim \frac{c_1}{(s_+ - s)^{c_2}}, \qquad s \uparrow s_+, \tag{3.3}$$

since the first integral in (3.3) is $O(1)$. By Karamata's Tauberian theorem [5, Theorem 1.7.1], we obtain from (3.2) that

$$\int_0^k e^{s_+ x} f(x,T)dx \sim \frac{c_1 k^{c_2}}{\Gamma(c_2+1)}, \qquad k \to \infty,$$

hence, by differentiating,

$$f(k,T) \approx e^{-s_+ k} \frac{c_1 k^{c_2 - 1}}{\Gamma(c_2)}, \qquad k \to \infty. \tag{3.4}$$

Similarly, the asymptotics

$$\frac{\partial_T m(s,T)}{s(s-1)} M(s,T) \sim \frac{\dot{c}_2}{s_+(s_+ - 1)} \log \frac{1}{s_+ - s} \times \frac{c_1}{(s_+ - s)^{c_2}}, \qquad s \uparrow s_+,$$

imply that the numerator in (2.4) approximately equals

$$\approx \frac{2c_1\dot{c}_2}{s_+(s_+-1)\Gamma(c_2)}e^{-s_+k}k^{c_2-1}\log k. \tag{3.5}$$

Now divide (3.5) by (3.4) to see that $\sigma^2_{\mathrm{loc}}(k,T)$ approximately equals (3.1). Note that we did not talk about Tauberian conditions, which are necessary to make this derivation rigorous, such as monotonicity of the density. In concrete cases, where an analytic continuation of the mgf is available, a Hankel contour approach [13] might be preferable to Karamata's theorem.

4 Local vol at extreme strikes in the Heston model

In this section we give a numerical example and explain how (1.3) is obtained by specializing (1.6). (But recall that (1.6) is so far just a recipe and not a theorem; a rigorous proof of (1.3) is given in Appendix A.)

Figure 1 compares the approximations (1.3) and (1.6) for the local vol. While asymptotically equivalent, the plot suggests that (1.3) has an $O(1)$ error term, whereas the error of (1.6) seems to be only $o(1)$. Note that the right hand side of (1.6) can be easily evaluated numerically, by using the explicit expression [18] of the Heston mgf in (1.7). We will now show that the right hand side of (1.6) is

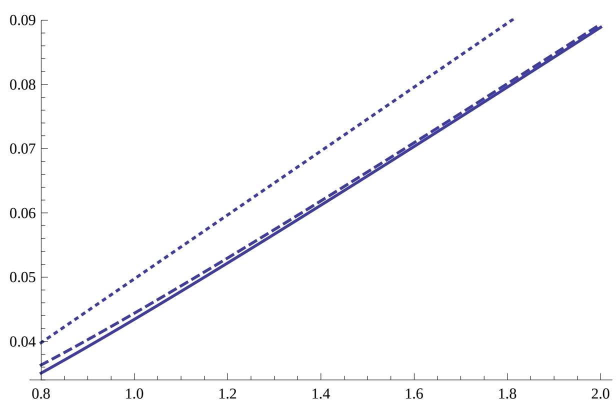

Figure 1: Local volatility squared $\sigma^2_{\mathrm{loc}}(k,T)$ (solid curve) in the Heston model, with parameters $T=1$, $a=0.0428937$, $b=-0.6067$, $c=0.2928$, $s_0=1$, $v_0=0.0654$, $\rho=-0.7571$. The approximation (1.6) is dashed, and (1.3) is dotted.

indeed asymptotically equivalent to the right hand side of (1.3). This requires us to know that

$$2\left.\frac{\partial}{\partial T}m(s,T)\right|_{s=\hat{s}(k,T)} \sim 2k/\sigma, \tag{4.1}$$

where $\sigma = \sigma(T)$ is the so-called critical slope, defined as

$$\begin{aligned} \sigma(T) &= -\frac{\partial T^*}{\partial s}(s_+(T)), \\ T^*(s) &= \sup\{t \geq 0 : \mathbb{E}[e^{sX_t}] < \infty\}. \end{aligned} \tag{4.2}$$

In fact, while the computation of the critical exponent s_+ in the Heston model requires simple numerics, the critical slope can be computed in closed form [14]; we have $\sigma(T) = TR_1/R_2$, where $R_i = R_i(b, c, \rho, s_+(T))$, $i = 1, 2$, are defined in (1.4)–(1.5).

Since $\hat{s}(k, T) \to s_+(T)$ as $k \to \infty$, the right hand side of (1.6) then satisfies[2]

$$\left.\frac{2\frac{\partial}{\partial T}m(s,T)}{s(s-1)}\right|_{s=\hat{s}(k,T)} \sim \frac{2}{\sigma(T)s_+(T)(s_+(T)-1)} \times k, \qquad k \to \infty,$$

which is the formula from Theorem 1. Let us now discuss validity of (4.1). The argument which follows nicely illustrates how formula (1.6) is used in stochastic volatility models of affine type. First, $m(s, t) \approx v_0\psi(s, t)$ for a function ψ for which we know[3]

$$\psi(s,t) \sim \frac{1}{\frac{c^2}{2}(T^*(s)-t)}, \qquad t \uparrow T^*(s),$$

and also

$$\frac{\partial}{\partial t}\psi(s,t) \sim \frac{1}{\frac{c^2}{2}(T^*(s)-t)^2}, \qquad t \uparrow T^*(s).$$

If we write $s_+ = s_+(T)$ when T is fixed, this translates to

$$m(s,t) \sim \frac{v_0}{\frac{c^2}{2}\sigma(s_+-s)}, \qquad s \uparrow s_+,$$

$$\frac{\partial}{\partial s}m(s,t) \sim \frac{v_0}{\frac{c^2}{2}\sigma(s_+-s)^2}, \qquad s \uparrow s_+, \tag{4.3}$$

$$\frac{\partial}{\partial T}m(s,T) \sim \frac{v_0}{\frac{c^2}{2}(\sigma(s_+-s))^2}, \qquad s \uparrow s_+. \tag{4.4}$$

Equation (1.7) leads to $\hat{s} = s_+ - \beta k^{-1/2} + o(k^{-1/2})$, since

$$\frac{\partial}{\partial s}m(s,t) \sim \frac{v_0}{\frac{c^2}{2}\sigma(s_+-\hat{s})^2} = k \implies s_+ - \hat{s} \sim \beta k^{-1/2}$$

with $\beta = \frac{\sqrt{2v_0}}{c\sqrt{\sigma}}$. Substitution then yields

$$\frac{\partial}{\partial T}m(s,T)|_{s=\hat{s}} \sim \frac{v_0}{\frac{c^2}{2}\sigma^2\beta^2/k} = k/\sigma,$$

which concludes our derivation of (4.1).

[2]It is worth noting that $\sigma(T) \sim \mathit{const} \times T$ as $T \to \infty$. This suggests that $\sigma^2_{\text{loc}}(kT, T)$ admits a non-degenerate limit as $T \to \infty$; Gatheral's SVI limit of Heston implied volatility was obtained in a similar regime.

[3]This follows from a straightforward analysis of the Riccati equations [14].

5 Some remarks on Dupire's formula for jump models

As discussed in the introduction, a direct application of Dupire's formula is not easy to justify in the presence of jumps. Even so, given the industry practice of applying Dupire's formula to any given call price surface, we now discuss what happens when applying formula (1.6) to some examples of jump models.

Example 2 (Double exponential Lévy)**.** *For zero drift, the mgf is given by [8]*

$$M(s,T) = \exp\left(T\left(\frac{\sigma^2 s^2}{2} + \lambda\left(\frac{\lambda_+ p}{\lambda_+ - s} + \frac{\lambda_-(1-p)}{\lambda_- + s}\right)\right)\right).$$

The critical moment is $s_+ = \lambda_+$, and the saddle point is located at

$$\hat{s} \approx s_+ - \sqrt{\frac{\lambda\lambda_+ pT}{k}}. \tag{5.1}$$

Formula (1.6) *thus yields*

$$\sigma^2_{\mathrm{loc}}(k,T) \approx \frac{2\sqrt{\lambda p}}{\sqrt{\lambda_+ T}(\lambda_+ - 1)} k^{1/2}. \tag{5.2}$$

In Figure 2, the fit of (5.2) *is not satisfactory (the dotted curve). Similarly to the Heston model, the approximation* (5.2) *has on $O(1)$ error term, whereas* (1.6) *seems to have $o(1)$, and gives a very good estimate.*

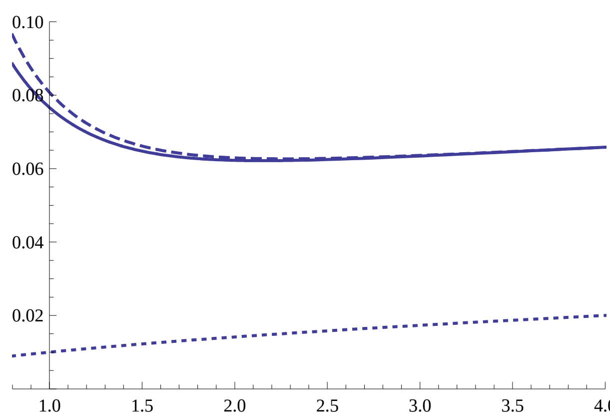

Figure 2: Local volatility squared $\sigma^2_{\mathrm{loc}}(k,T)$ (solid curve) in the double exponential Lévy model, with parameters $T = 1$, $\sigma = 0.2$, $\lambda = 10$, $p = 0.3$, $\lambda_- = 25$, $\lambda_+ = 50$. The approximation (1.6) is dashed, whereas (5.2) is dotted.

As mentioned in Section 1, one expects local volatility to explode for $T \downarrow 0$: We have $\sigma^2_{\mathrm{loc}}(k,T) \approx$ const $\times\, T^{-1/2}$ in the double exponential Lévy model.

To see this, note that in Lévy models the saddle point $\hat{s}(k,T)$ is a function of k/T, and that the saddle point method works for $T \downarrow 0$ as well as for $k \to \infty$. Therefore, (5.2) *is true for fixed k and $T \downarrow 0$, too.*

Example 3 (Variance gamma)**.** *The mgf is given by [21]*

$$M(s,T) = \left(\frac{1}{1 - \theta\nu s - \frac{1}{2}\sigma^2 \nu s^2} \right)^{T/\nu}.$$

We assume that $T > \nu/2$, which guarantees that the log-price has a density, and hence that call prices are C^2 (see Example 1 in [9]). The critical moment is

$$s_+ = \frac{\sqrt{2\nu\sigma^2 + \nu^2\theta^2} - \nu\theta}{\nu\sigma^2}.$$

Since we have

$$m(s,T) = \log M(s,T) \sim \frac{T}{\nu} \log \frac{1}{s_+ - s},$$

the saddle point satisfies

$$\hat{s} \approx s_+ - \frac{T}{\nu k}.$$

By (1.6)*, we thus have*

$$\sigma^2_{\mathrm{loc}}(k,T) \approx \frac{2\log(k/T)}{\nu s_+(s_+ - 1)}. \tag{5.3}$$

According to Figure 3, this approximation kicks in only for fairly large values of k. As in Section 4 and Example 2, an improved estimate is obtained by using (1.6) *directly.*

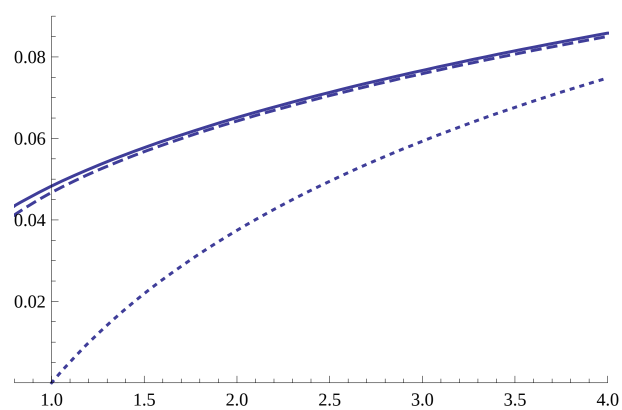

Figure 3: Local volatility squared $\sigma^2_{\mathrm{loc}}(k,T)$ (solid curve) in the variance gamma model, with parameters $T = 1$, $\sigma = 0.261652$, $\theta = -0.218033$, $\nu = 0.0552584$. The approximation (1.6) is dashed, and (5.3) is dotted.

Example 4 (Normal inverse Gaussian). *This is an example where condition* (1.2) *is violated, and our formula* (1.6) *does not hold. The mgf*

$$M(s,T) = \exp\left(\delta T\left(\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+s)^2}\right)\right)$$

has no blow-up at the critical moment

$$s_+ = \alpha - \beta,$$

but a square-root type singularity, with local expansion

$$M(s,T) \approx e^{\delta T\sqrt{\alpha^2-\beta^2}}\left(1 - \delta T\sqrt{2\alpha}\sqrt{s_+ - s}\right). \tag{5.4}$$

It is still true that $\sigma^2_{\mathrm{loc}}(k,T)$ *asymptotically depends, via* (2.4)*, on the local behavior of* $M(s,T)$ *near* s_+*. However, the approximation* (1.6) *hinges on the* first *term of the local expansion of* $M(s,T)$*. It therefore fails to capture the asymptotics of* $\sigma^2_{\mathrm{loc}}(k,T)$*, which depend on the first* singular *term (the term* $\sqrt{s_+ - s}$ *in* (5.4)*). An analysis can be done by a Hankel contour approach [13], and yields*

$$\sigma^2_{\mathrm{loc}}(k,T) \approx \frac{2\left(1+\delta T\sqrt{\alpha^2-\beta^2}\right)}{s_+(s_+-1)}.$$

The numerical fit is not very good, though, and further terms should be computed for improved accuracy. We propose to return to this model and the more general GH (generalized hyperbolic) model in a future study.

The fact that $\sigma^2_{\mathrm{loc}}(k,T)$ *converges to a constant might be understood by comparing the NIG marginals with those of Heston's in the time* $T\to\infty$ *regime (this link is made precise in [19]). In particular, the result is then consistent with the Heston asymptotics* (1.3) *of local vol; note that the right hand side of* (1.3) *is* $O(1/T)$ *for* $T\to\infty$*.*

6 Conclusions

We propose a new formula that expresses local volatility for extreme strikes as a computable function of commonly available model information. In the Heston model this leads to a proof that local volatility (squared) behaves asymptotically linear in log-strike (which is qualitatively similar to Lee's result [20] for implied volatility).

Although we suspect that this Lee-type behavior remains true for models similar enough to Heston (e.g. local stochastic volatility models with a Heston backbone [17]), qualitatively different behavior is seen in models with jumps. We derived this by applying our generic approximation formula (1.6), supported by numerical examples.

While this enhances our knowledge of local volatility in a variety of models, it also has a clear impact on calibration of local volatility to market data. Indeed,

liquid option data is typically available only in a restricted range of strikes and maturities; our results can then be used to extrapolate local volatility in a way that is consistent with Heston stochastic volatility or other chosen models. In particular, this approach avoids any arbitrage possibilities introduced by ad-hoc specifications of the implied volatility surface. We also believe the present methodology will turn out useful in the calibration of local stochastic volatility models to market smiles.

Acknowledgment. We thank Rama Cont and Jim Gatheral for helpful discussions.

A Proof of Theorem 1 (local vol approximation for the Heston model)

By the exponential decay of the Heston mgf towards $\pm i\infty$, the formulas (2.1)–(2.4) are correct for the Heston model. For the saddle point analysis of (2.4), we employ the approximate saddle point

$$\hat{s}_{\mathrm{approx}}(k) := s_+ - \beta k^{-1/2},$$

obtained by using (4.3) in (1.7). (Recall that $\beta = \sqrt{2v_0}/c\sqrt{\sigma}$, and that σ denotes the critical slope defined in (4.2).) This approximate saddle may be used for both integrals in (2.4). As for the denominator, this was carried out in detail in [14], where an expansion of the Heston density was determined. The analysis of the numerator in (2.4) is similar, except that a new tail estimate is required. But first we discuss the local expansion around the saddle point. Let us fix a number $\alpha \in (\frac{2}{3}, \frac{3}{4})$ and define $h(k) = k^{-\alpha}$. Then, in the central range $|s - \hat{s}_{\mathrm{approx}}(k)| \leq h(k)$, we have

$$\begin{aligned}\frac{1}{s(s-1)} &= \frac{1}{s_+(s_+-1)} + O(s_+ - s) \\ &= \frac{1}{s_+(s_+-1)}\Big(1 + O(k^{-1/2})\Big)\end{aligned}$$

and (cf. (4.4))

$$\begin{aligned}2\frac{\partial}{\partial T} m(s,T) &= \frac{2\beta^2}{\sigma(s_+-s)^2} + O\left(\frac{1}{s_+-s}\right) \\ &= \frac{2\beta^2}{\sigma}(\beta k^{-1/2} + O(k^{-\alpha}))^{-2} + O(k^{-1/2}) \\ &= \frac{2k}{\sigma}(1 + O(k^{1/2-\alpha})).\end{aligned}$$

Therefore, the local expansions of the two integrands in (2.4) agree, up to a factor that is given by

$$\frac{2\partial_T m(s,T)}{s(s-1)} = \frac{2k}{\sigma s_+(s_+-1)}(1 + O(k^{1/2-\alpha})), \tag{A.1}$$

where the error term holds uniformly w.r.t. the integration variable s. According to Theorem 1.2 of [14], we have

$$\frac{1}{2i\pi}\int_{\hat{s}_{\text{approx}}-ih(k)}^{\hat{s}_{\text{approx}}+ih(k)} e^{-ks}M(s,T)ds \sim A_1 e^{(1-A_3)k+A_2\sqrt{k}}k^{-3/4+a/c^2} \tag{A.2}$$

for certain constants A_1, $A_2 = 2\beta$, and $A_3 = s_+ + 1$. Analogously, we derive from (A.1) that

$$\frac{1}{2i\pi}\int_{\hat{s}_{\text{approx}}-ih(k)}^{\hat{s}_{\text{approx}}+ih(k)} \frac{2\partial_T m(s,T)}{s(s-1)} e^{-ks}M(s,T)ds \\ \sim \frac{2k}{\sigma s_+(s_+-1)} \times A_1 e^{(1-A_3)k+A_2\sqrt{k}}k^{-3/4+a/c^2}. \tag{A.3}$$

Dividing (A.3) by (A.2) shows our claim (1.3), provided that the tails $|s - \hat{s}_{\text{approx}}(k)| > h(k)$ of the integrals can be discarded. For the denominator of (2.4), this was shown in Lemma A.3 of [14]. So we proceed with the numerator. We consider only the upper tail, as the lower one is handled by symmetry. By Lemma A.3 of [14], there is a constant $B > 0$ such that

$$\left|\int_{\hat{s}_{\text{approx}}+ih(k)}^{\hat{s}_{\text{approx}}+iB} e^{-ks}M(s,T)ds\right| \le e^{(1-A_3)k}\exp(A_2\sqrt{k}-\tfrac{1}{2}\beta^{-1}k^{3/2-2\alpha}+O(\log k)). \tag{A.4}$$

From (4.3) we obtain

$$\left|\frac{\partial_T m(s,T)}{s(s-1)}\right| \le const \times k$$

for all s on the contour in (A.4). This estimate can be absorbed into the factor $\exp(O(\log k))$ in (A.4), so that we conclude

$$\left|\int_{\hat{s}_{\text{approx}}+Ih(k)}^{\hat{s}_{\text{approx}}+iB} \frac{\partial_T m(s,T)}{s(s-1)} e^{-ks}M(s,T)ds\right| \\ \le e^{(1-A_3)k}\exp(A_2\sqrt{k}-\tfrac{1}{2}\beta^{-1}k^{3/2-2\alpha}+O(\log k)). \tag{A.5}$$

This grows slower than (A.3) (compare the relevant factors $k^{-3/4+a/c^2}$ resp. $\exp(-\frac{1}{2}\beta^{-1}k^{3/2-2\alpha})$). As for $\Im(s) > B$, it was shown in [14] (Lemma A.2) that

$$\left|\int_{\hat{s}_{\text{approx}}+iB}^{\hat{s}_{\text{approx}}+i\infty} e^{-ks}M(s,T)ds\right| = O(\exp((1-A_3)k+\beta\sqrt{k})).$$

This was deduced from the exponential decay of $M(s,T)$ for large $\Im(s)$ (Lemma A.1 in [14]). The following lemma implies that the new factor $\partial_T m(s,T)/(s(s-1))$ grows only polynomially, so that the exponential decay of the integrand persists for the numerator of (2.4). This finishes the proof of Theorem 1.

To state the lemma, recall that $m(s,t) = \phi(s,t) + v_0\psi(s,t)$, where ϕ and ψ satisfy the Riccati equations

$$\begin{aligned}
&\dot{\phi} = a\psi, \qquad \phi(0) = 0,\\
&\dot{\psi} = \tfrac{1}{2}(s^2 - s) + \tfrac{1}{2}c^2\psi^2 + b\psi + s\rho c\psi, \qquad \psi(0) = 0.
\end{aligned}$$

We have to show that $\dot{m}$ grows only polynomially as $\Im(s) \to \infty$. Because of the Riccati equations, it suffices to show this for ψ. Let us write $\psi = f + ig$ and $s = \xi + iy$.

Lemma 5. *Let $T > 0$, and assume that the real part ξ of s stays bounded in some interval $1 \le \xi \le \xi_{\max}$. Then, there are positive constants $C_{i,T}$ $(i = 1,2,3,4)$ such that for $y \ge y_0$, where y_0 depends only on $\xi_{\max}$ and the other (fixed) model parameters of the Heston model,*

$$\begin{aligned}
-C_{3,T}y^2 &\le f(t) \le -C_{1,T}y,\\
-C_{4,T}y^3 &\le g(t) \le C_{2,T}\,y.
\end{aligned}$$

In fact, we can take

$$\begin{aligned}
C_{1,T} &= 1/\left(3c\right),\\
C_{2,T} &= \frac{1}{2}\left(2\xi_{\max} - 1\right)T,\\
C_{3,T} &= T\left(1 + \frac{c^2}{2}C_{2,T}^2\right),\\
C_{4,T} &= 2C_{3,T}Tc^2C_{2,T}.
\end{aligned}$$

Proof. It follows from the proof of Lemma A.1 in [14] that (e.g. with $C_{1,T} := T\theta = \frac{1}{c}\sqrt{1/6} \le \frac{1}{3c}$)

$$f(t) \le -T\theta y = -\frac{1}{c}\sqrt{1/6}y \le -\frac{1}{3c}y =: -C_{1,T}y.$$

We next provide a similar upper estimate for g. To this end we first show that $g = g(t)$ remains ≥ 0 for all times $t > 0$. The differential equation for g,

$$\dot{g} = \frac{1}{2}\left(2\xi y - y\right) + c^2 fg - \gamma g, \qquad g(0) = 0,$$

implies the first order Euler estimate

$$\begin{aligned}
g\left(t\right) &= g(0) + \left\{\frac{1}{2}\left(2\xi y - y\right) + c^2 f(0)g(0) - \gamma g(0)\right\}t + o(t)\\
&= \underbrace{\frac{1}{2}\left(2\xi y - y\right)}_{>0}t + o(t),
\end{aligned}$$

and hence g is positive (even strictly so) on some interval $(0, \varepsilon_1)$. Assume this interval is maximal in the sense that $g(\varepsilon_1) = 0$ and g is (strictly) negative on

some further interval $(\varepsilon_1, \varepsilon_2)$. Clearly then $\dot{g}(\varepsilon_1) \leq 0$, which contradicts the information from the differential equation: indeed, using $g(\varepsilon_1) = 0$, we obtain the contradiction

$$\dot{g}(\varepsilon_1) = \underbrace{\frac{1}{2}(2\xi y - y)}_{>0}.$$

The observation that $g \geq 0$ is useful to us, since it leads, together with $f \leq -C_{1,T} y$ and $\gamma \geq 0$, to the differential inequality

$$\begin{aligned}
\dot{g} &= \frac{1}{2}(2\xi y - y) + c^2 f g - \gamma g \\
&\leq \frac{1}{2}(2\xi y - y) - \left(c^2 C_{1,T} + \gamma\right) g \\
&\leq \frac{1}{2}(2\xi y - y),
\end{aligned}$$

and hence to the upper estimate

$$\forall 0 \leq t \leq T : g(t) \leq \frac{1}{2}(2\xi_{\max} - 1)\, T \times y =: C_{2,T} y.$$

We can feed this upper estimate on g back in the differential equation for f to obtain a lower estimate

$$\begin{aligned}
\dot{f} &= \frac{1}{2}\left(\xi^2 - y^2 - \xi\right) + \frac{c^2}{2}\left(f^2 - g^2\right) - \gamma f \\
&\geq \frac{1}{2}\left(\xi^2 - y^2 - \xi\right) + \frac{c^2}{2} f^2 - \frac{c^2}{2} C_{2,T}^2 y^2 - \gamma f \\
&= -\frac{1}{2}\left(1 + c^2 C_{2,T}^2\right) y^2 + \frac{1}{2}\left(\xi^2 - \xi\right) - \gamma f + \frac{c^2}{2} f^2 \\
&\geq -\frac{1}{2}\left(1 + c^2 C_{2,T}^2\right) y^2 + \frac{1}{2}\left(\xi^2 - \xi\right) - \gamma f \\
&\geq -\left(1 + \frac{c^2}{2} C_{2,T}^2\right) y^2 - \gamma f,
\end{aligned}$$

where in the last step we assume that y is large enough so that the extra amount subtracted (at least: $\frac{1}{2} y^2$) is larger than $\frac{1}{2}\left(\xi^2 - \xi\right)$, which remains bounded. We also know that $f(t) \leq -C_{1,T} y \leq 0$ for all $0 \leq t \leq T$. It follows that $-\gamma f \geq 0$ and omission leads to our final lower bound on $\dot{f}$, namely

$$\dot{f} \geq -\left(1 + \frac{c^2}{2} C_{2,T}^2\right) y^2.$$

This entails immediately

$$f(t) \geq -T\left(1 + \frac{c^2}{2} C_{2,T}^2\right) y^2 =: -C_{3,T} y^2.$$

At last, we need a lower bound on g. Again, we look for a suitable differential inequality. Since $g \geq 0$,

$$\begin{aligned}\dot{g} &= \frac{1}{2}\left(2\xi y - y\right) + c^2 f g - \gamma g \\ &\geq \frac{1}{2}\left(2\xi y - y\right) - c^2 \left|f\right| g - \gamma g \\ &\geq \frac{1}{2}\left(2\xi - 1\right) y - \left(C_{3,T} y^2 c^2 + \gamma\right) g \\ &\geq \frac{1}{2}\left(2\xi - 1\right) y - 2C_{3,T} y^2 c^2 g,\end{aligned}$$

for y large enough such that the additional subtraction of $C_{3,T}y^2c^2$ takes care of γ. Using the upper estimate on g (linear in y), and the fact that $(2\xi - 1) \geq 0$, we conclude

$$\dot{g} \geq -2C_{3,T}y^2c^2C_{2,T}y.$$

It immediately follows (since $g(0) = 0$) that

$$\forall 0 \leq t \leq T : g(t) \geq -2C_{3,T}Tc^2C_{2,T}y^3 =: -C_{4,T}y^3.$$

■

References

[1] L. B. G. Andersen and V. V. Piterbarg, *Moment explosions in stochastic volatility models*, Finance Stoch., 11 (2007), pp. 29–50.

[2] M. Avellaneda, D. Boyer-Olson, J. Busca, and P. K. Friz, *Reconstruction of volatility: Pricing index options using the steepest-descent approximation*, Risk, (October 2002), pp. 91–95.

[3] A. Bentata and R. Cont, *Forward equations for option prices in semimartingale models.* Preprint, available at http://arxiv.org/abs/1001.1380, 2010.

[4] H. Berestycki, J. Busca, and I. Florent, *Computing the implied volatility in stochastic volatility models*, Comm. Pure Appl. Math., 57 (2004), pp. 1352–1373.

[5] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular variation*, vol. 27 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 1987.

[6] P. Carr, H. Geman, D. P. Madan, and M. Yor, *From local volatility to local Lévy models*, Quant. Finance, 4 (2004), pp. 581–588.

[7] P. Carr and D. P. Madan, *Option valuation using the fast Fourier transform*, Journal of Computational Finance, 3 (1999), pp. 463–520.

[8] R. Cont and P. Tankov, *Financial modelling with jump processes*, Chapman & Hall/CRC Financial Mathematics Series, Chapman & Hall/CRC, Boca Raton, FL, 2004.

[9] R. Cont and E. Voltchkova, *Integro-differential equations for option prices in exponential Lévy models*, Finance Stoch., 9 (2005), pp. 299–325.

[10] N. G. de Bruijn, *Asymptotic methods in analysis*, Bibliotheca Mathematica. Vol. 4, North-Holland Publishing Co., Amsterdam, 1958.

[11] B. Dupire, *Pricing with a smile*, Risk, 7 (1994), pp. 18–20.

[12] ——, *A unified theory of volatility.* Working paper, Paribas, 1996.

[13] W. B. Ford, *Studies on divergent series and summability and the asymptotic developments of functions defined by Maclaurin series*, Chelsea Publishing Company, 3rd ed., 1960. (From two books originally published in 1916 and 1936.).

[14] P. Friz, S. Gerhold, A. Gulisashvili, and S. Sturm, *On refined volatility smile expansion in the Heston model.* To appear in Quantitative Finance, 2011.

[15] J. Gatheral, *The Volatility Surface, A Practitioner's Guide*, Wiley, 2006.

[16] P. Henry-Labordere, *A general asymptotic implied volatility for stochastic volatility models.* Available at SSRN: http://ssrn.com/abstract=698601, 2005.

[17] P. Henry-Labordere, *Calibration of local stochastic volatility models to market smiles: A Monte-Carlo approach*, Risk, (September 2009). Available at SSRN: http://ssrn.com/abstract=1493306.

[18] S. Heston, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, Review of Financial Studies, 6 (1993), pp. 327–343.

[19] M. Keller-Ressel, *Moment explosions and long-term behavior of affine stochastic volatility models*, Math. Finance, 21 (2011), pp. 73–98.

[20] R. W. Lee, *The moment formula for implied volatility at extreme strikes*, Math. Finance, 14 (2004), pp. 469–480.

[21] D. Madan, P. Carr, and E. Chang, *The variance gamma process and option pricing*, European Finance Review, 2 (1998), pp. 79–105.

[22] V. Piterbarg, *Markovian projection method for volatility calibration.* Available at SSRN: http://ssrn.com/abstract=906473, 2006.